

УДК 519.866

МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ЦІНИ НА РИНКУ НАФТИ ЗАСОБАМИ ARIMA-МОДЕЛІ

Запорожчук Тетяна

Науковий керівник: канд.ф.-м. наук Гаєвський М.В.

Центральноукраїнський державний педагогічний університет імені

Володимира Винниченка, м. Кропивницький, Україна

В статті розглянуто застосування лінійних стохастичних моделей для дослідження ринку нафти. Зокрема, побудовано та досліджено різні ARIMA моделі ціни нафти та на основі їх зроблено прогноз ціни.

Ключові слова: нафта, часовий ряд, ARIMA-модель, авторегресія, змінне (ковзне) середнє.

Modeling and forecasting prices of oil by ARIMA models

T. Zaporozhchuk

Scientific supervisor: Candidate of Physics and Mathematics Sciences Haievskiy M.V.

The Volodymyr Vynnychenko Central Ukrainian State Pedagogical University,

Kropyvnytsky, Ukraine

In the article the application of linear stochastic models for investigating the oil market is considered. In particular, various ARIMA oil price models were constructed and investigated and based on their price forecast.

Key words: oil, time series, ARIMA-model, autoregression, moving average.

Постановка проблеми. В сучасній статистичній теорії існує багато різноманітних методів прогнозування інформації. Значна їх частина відноситься до прогнозування часових рядів. Особливістю прогнозування часових рядів є те, що аналізуються лише дані спостережень без додаткової інформації, без аналізу впливу зовнішніх сил. Звичайно, такий аналіз виглядає досить неповним, але доволі часто прогнози часових рядів є дуже. Більш детально із методикою дослідження часових рядів та методами їх прогнозування можна ознайомитися, наприклад, у роботах [1,2,4,7-15].

Також слід відмітити, що більшість економічних, соціологічних процесів являють собою нестационарні часові ряди, а це в свою чергу створює певні проблеми для їх дослідження. Одним із поширених підходів для дослідження

таких процесів є методика Бокса-Дженкінса.

Підхід Бокса-Дженкінса до аналізу часових рядів є дуже потужним інструментом для побудови точних прогнозів. Вперше систематичний підхід до побудови моделі ARIMA був викладений Боксом і Дженкінсом. Моделі ARIMA є досить гнучкими та можуть описувати широкий спектр характеристик часових рядів, що зустрічаються на практиці та мають багато практичних застосувань. Однією із головних властивостей ARIMA-моделей є можливість їх використання для нестационарних часових рядів.

Такі моделі включають в себе авторегресію та змінне (ковзне) середнє. У моделі авторегресії кожне значення ряду знаходиться в лінійній залежності від попередніх значень. Модель змінного середнього передбачає, що в помилках моделі в попередні періоди зосереджена інформація про всю передісторію ряду.

Для будь-якого часового ряду можна створити деяку модель, яка б відображала розвиток процесу. Але яка б не була модель, неможливо прогнозувати значення часового ряду без похибки. Таким чином, будь-яка модель повинна містити похибку, за величиною якої вимірюється точність моделі. Цією похибкою якраз і є випадкова складова. Для стаціонарних часових рядів випадкова складова будується з допомогою такого часового ряду, який прийнято називати білим шумом.

Аналіз досліджень і публікацій. В роботі [3] досліджується ціна на нафту – фактори впливу на формування її ціни, вплив її на світову економіку та важливість побудови якісних прогнозів, в [5] досліджуються сучасні тренди науково-технологічного розвитку та їх роль у нафтовому секторі. Робота [6] присвячена ринку нафти та нафтопродуктів України, проблемам його розвитку та регулювання.

Мета статті. Метою нашої роботи є моделювання та прогнозування ціни на ринку нафти засобами ARIMA.

Вперше систематичний підхід до побудови моделі ARIMA був викладений Боксом і Дженкінсом в 1976 році [2]. Побудова ARIMA-моделі для досліджуваних часових рядів включає наступні основні етапи [2, 11]:

- Ідентифікацію пробної моделі;
- Оцінювання параметрів моделі та діагностичну перевірку адекватності моделі;
- Використання моделі для прогнозування.

Структурна схема вибору моделі ARIMA

1. Ряд даних
2. Перевірка на стаціонарність
3. Перетворення ряду шляхом послідовного взяття різниць
4. Ідентифікація базового набору моделей
5. Оцінювання параметрів моделі
6. Діагностична перевірка на адекватність і вибір моделі
7. Використання моделі для прогнозування

Таким чином, спочатку (в блоці 1-3) необхідно отримати стаціонарний ряд. На цьому етапі рекомендується проводити аналіз автокореляційної функції (АКФ) та приватної автокореляційної функції (ЧАКФ). Швидке загасання значень АКФ - простий тест на стаціонарність. На цьому етапі використовуються також статистичні тести на наявність одиничного кореня (розширений тест Дікі-Фуллера або ADF-тест) .

Якщо відповідно до статистики Дікі-Фуллера або оцінок АКФ ряд є нестаціонарним, то для переходу до стаціонарного ARIMA-модель прогнозування значень ряду традиційно застосовують оператор взяття послідовних різниць, тим самим виділяється значення параметра **d** (близька різниця). Таким чином, значення одного параметра в моделі ARIMA (p, d, q) вже відомо.

У блоці 4 після отримання стаціонарного ряду досліджується характер поведінки вибірових АКФ і ЧАКФ і висуваються гіпотези про значення параметрів **p** (порядок авторегресії) і **q** (порядок змінного середнього). На вході блоці 4 може формуватися базовий набір, включаючи одну, дві або навіть більше число моделей, іншими словами, портфель моделей.

У блоці 5 після здійснення ідентифікації моделі необхідно оцінити їх

параметри. Для цих цілей використовується метод максимальної правдоподібності (ММП).

У блоці 6 для перевірки кожної пробної моделі на адекватність аналізується її ряд залишків. У адекватної моделі ряд залишків повинен бути схожий на білий шум, тобто їх вибіркові АКФ не повинні відрізнятися від нуля. Для перевірки гіпотези про те, що спостерігаємо дані є реалізацією «білого шуму», використовується також Q-статистика Льюїнга-Бокса:

$$Q = n + 2 \sum_{k=1}^m \frac{r_k^2}{n-k},$$

де n - обсяг вибірки, m - максимальна кількість лагів, r_k - коефіцієнти автокореляційної функції.

Якщо в результаті перевірки кілька моделей виявляються адекватними, то при остаточному виборі слід врахувати два фактори:

- Підвищення точності (якість підгонки моделі);
- Зменшення числа параметрів моделі. Спільно ці вимоги зведені в інформаційні критерії Акайке і Шварца.

За допомогою моделі в блоці 7 можна будувати точний і інтервальний прогноз на L кроків вперед.

Для оцінки точності прогнозу використовується ряд стандартних показників: абсолютні показники – середньоквадратична похибка MSE, корінь із середньоквадратичної похибки RMSE, середня абсолютна похибка MAD; відносні показники – корінь із середньоквадратичної похибки у відсотках від фактичних значень за n кроків RMSPE, середня абсолютна похибка у відсотках за n кроків MAPE

На практиці для процесів ARMA високих порядків досить часто буває, що модель низького порядку має менше параметрів та екстраполює дані майже так само, як і модель вищого порядку [8-9].

При прийнятті рішень потрібно враховувати оцінювання ризику, що є основним на ринку. Так як, більшість ризиків і винагород знаходяться у майбутньому, інвестори змушені враховувати всю загальну інформацію, а саме:

очікування втрат і винагород, приймаючи рішення, що покращують інвестиції, максимізуючи винагороди та мінімізуючи ризики. Мірою ризику визначають зазвичай ринкову нестабільність або діапазон коливань ціни активу. Визначивши можливий діапазон коливань курсу акцій, інвестор може адекватно оцінювати свої шанси на виграш і вже тоді приймати ефективніші рішення. На фінансових ринках важливим є випадкові відхилення величини від постійного значення впродовж часу, оскільки вартість акцій, опціонів та інших фінансових інструментів залежить від рівня їх ризикованості. В цьому випадку дисперсію (волатильність) прибутковості природно розглядати як міру ризикованості фінансового активу. Тому, при прийнятті інвестиційних рішень корисно будувати прогнози не лише математичного сподівання шуканої величини, але і її дисперсії для наглядного бачення майбутньої ситуації.

Розглянемо максимальні щоденні дані цін у гривнях на ринку нафти за чотири місяці за допомогою них потрібно побудувати прогноз на 9 днів.

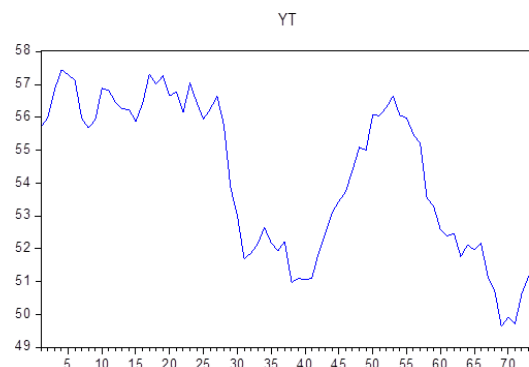
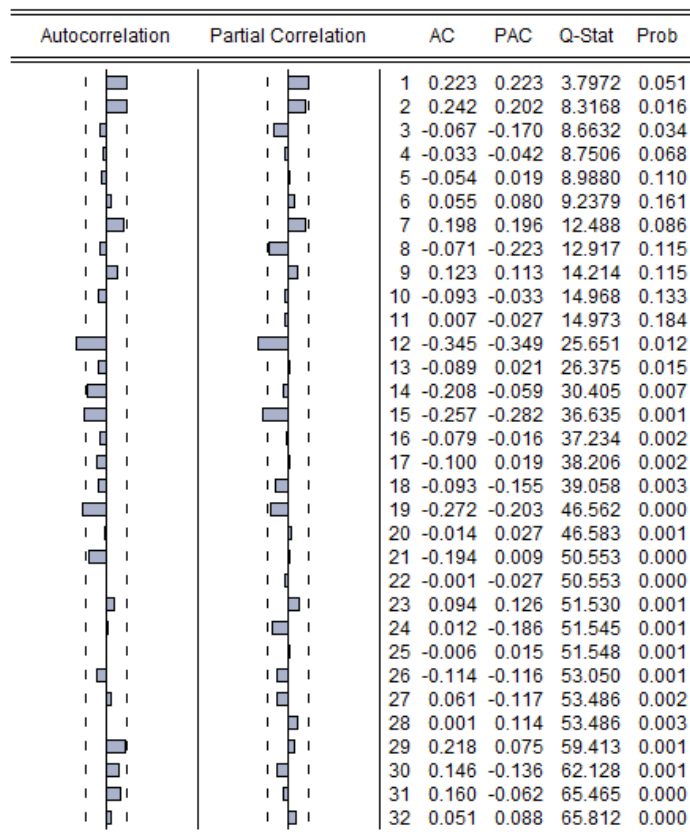


Рис 1. Графічна візуалізація цін на ринку нафти

Найпоширеніший тест перевірки часового ряду на стаціонарність — розширений тест Дікі–Фулера: Основний результат Дікі–Фулера тесту — отримання розрахункового значення та критичних значень t -статистики МакКінона. Наш ряд в рівнях нестационарний. Потрібно перевірити, чи зможемо ми операцією перших різниць перетворити ряд на стаціонарний. Ряд в перших різницях є стаціонарним, а ряд у рівнях **має порядок інтеграції 1**.

Отже, ARMA модель будуватимемо для ряду в перших різницях.

Проаналізувавши корелограму часового ряду в перших різницях, також відмітимо, що сезонність чітко не проявлена.



Для попереднього визначення загального вигляду специфікації майбутньої ARIMA моделі й кількості лагів для кожної складової скористаємося графіками автокореляційної і часткової автокореляційної функції досліджуваного показника. Візуальний аналіз ACF/PACF дає змогу зробити висновок, чи можна вважати часовий ряд чистим AR- або MA-процесом, чи він є змішаним ARMA процесом.

У випадку, коли таких лагів-кандидатів кілька або є сумніви щодо остаточного вибору лагу, можна додатково застосувати таку процедуру:

- 1) першою оцінити регресію, в яку включити максимально можливу кількість лагів. Проаналізувати значення AIC-критерію для цієї моделі;
- 2) видалити з оціненої регресії лаг із найбільшим значенням p-value для відповідного коефіцієнта і порівняти отримане значення AIC критерію з попереднім значенням;
- 3) якщо AIC став меншим, повторити процедуру з наступним лагом. Якщо значення AIC критерію зросло, повернути видалену на попередньому кроці лагову змінну в рівняння і видалити змінну з лагом, для якого p-value є наступним за величиною;

4) продовжувати процедуру до моменту, коли будь-який видалений лаг збільшуватиме значення AIC.

Початкове рівняння може містити не одну лагову змінну, а декілька, зокрема це можуть бути статистично значущі в графіках автокорелограми лагові змінні.

Аналіз показує, що в цьому випадку значущим є лаги 1, 2, 12, 15 та 19. Крім того, аналіз поведінки ACF/PACF свідчить про змішаний характер процесу.

Після визначення оптимальної AR-складової необхідно утворити ряд залишків цієї моделі для наступного використання при визначенні оптимального порядку MA-складової моделі ARMA/ARIMA.

Процедура пошуку оптимального порядку MA-складової моделі ARMA/ARIMA аналогічна процедурі визначення оптимального порядку AR-складової. Однак на цьому етапі до визначеної AR-складової поступово додають MA(1), MA(2).... MA(q)-складові і розраховують значення Шварца- критерію. Модель, яка має найменше значення критерію Шварца, – претендент для подальшого аналізу.

Попереднє оцінювання всіх описаних вище моделей для визначення оптимального порядку AR- та MA-складових здійснювалася за допомогою методу найменших квадратів. Однак застосування цього методу оцінювання є некоректним у випадку ARIMA моделювання, оскільки він може давати зміщені оцінки невідомих параметрів.

Тому після того, як визначено попередню специфікацію (форму) моделі, її потрібно переоцінити нелінійним методом найменших квадратів (NLS).

Для того, щоб модель вважалась адекватною, достатньо виконання однієї умови – залишки оціненого фінального рівняння мають бути білим шумом. Для перевірки залишків на білий шум, можна використати корелограми і тест Дікі–Фулера.

Прогнозування на основі побудованої ARIMA моделі

Дата	Максимум ціни	ARIMA(19,1,12)
------	---------------	----------------

		нафти на ринку	
15.05.2017	1	51,06	51.415
16.05.2017	2	51,10	51.487
17.05.2017	3	51,87	52.193
18.05.2017	4	52,48	52.806
19.05.2017	5	53,10	52.899
20.05.2017	6	53,45	52.745
21.05.2017	7	53,74	52.492
22.05.2017	8	54,38	53.085
23.05.2017	9	55,09	53.145
24.05.2017	9	55,15	53.714

Для оцінювання прогнозної якості моделі необхідно проаналізувати значення розрахованих критеріїв прогнозної якості. Одним із найпоширеніших на практиці є критерій середньої абсолютної процентної похибки (MAPE – Mean Absolute Percent Error), яка показує середню абсолютну похибку прогнозу в відсотках. Для нашого прикладу значення $MAPE = 0,64 \%$, що свідчить про високу прогнозну якість моделі.

Однак слід мати на увазі, що більш складні моделі мають кращі апроксимуючі властивості і одночасно менш стійкі, що для цілей прогнозування є визначальним негативним фактором. Тому для вибору кращої моделі використовують так звані інформаційні критерії: Акаїке, Шварца та ін. Згідно з критерієм Акаїке, серед альтернативних моделей вибирається та модель, для якої мінімізується величина AIC.

Оскільки сучасні статистичні пакети дають змогу будувати моделі даного типу з різними параметрами, то порівняємо наш прогноз із наступними.

Порівняння прогнозів ARIMA(p,d,q)

Дата	Реальні дані	(19,1,12)	(0,1,0)	(1,1,0)	(2,1,0)	(1,1,1)
15.05	51,06	51.415	51.026	51.020	51.148	50.667
16.05	51,10	51.487	50.963	50.950	51.076	50.579
17.05	51,87	52.193	50.901	50.880	51.001	50.494
18.05	52,48	52.806	50.837	50.811	50.924	50.409

19.05	53,10	52.899	50.774	50.741	50.844	50.324
20.05	53,45	52.745	50.711	50.671	50.767	50.239
21.05	53,74	52.492	50.647	50.602	50.693	50.154
22.05	54,38	53.085	50.584	50.532	50.622	50.069
23.05	55,09	53.145	50.521	50.462	50.547	49.984

Як бачимо, що коли процес ідентифікації не був проведений, то отримані прогнози справедливі на дуже короткий період, а прогноз, що побудований на основі дослідження найкраще описує і моделює досліджуване явище.

Висновки та перспективи подальших пошуків у напрямі дослідження. Нами досліджено динаміку цін на ринку нафти, зокрема побудовано ARIMA-моделі із різними параметрами для ціни на нафту для подальшого порівняння та отримано прогноз ціни на ринку нафти на основі ARIMA-моделей. Зроблено порівняння прогнозних значень із реальними даними та з іншими прогнозами.

Список літератури

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. / Айвазян С.А., Мхитарян В.С. – М.: ЮНИТИ, 1998. – 1000 с.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление / Бокс Дж., Дженкинс Г. – М.: Мир, 1994 г. – 425 с.
3. Брагинский О.Б. Цены на нефть: прогноз, влияние на економику // Рос. хим. ж. (Ж. Рос. хим. об-ва им. Д.И. Менделеева), 2008, т. LII, No 6 – с. 25-36с.
4. Брюков В.Г. Как предсказать курс доллара. Эффективные методы прогнозирования с использованием Excel и EViews. – М.: КНОРУС; ЦИПСИР, 2011. – 272 с.
5. Бурлака В.Г. Сучасні тренди науково-технологічного розвитку та їх роль у нафтовому секторі / В. Г. Бурлака // Актуальні проблеми економіки. - 2013. - № 5. - С. 54-65
6. Дорожкіна М. Ринок нафти та нафтопродуктів України: проблеми розвитку та регулювання / Вісник Київського національного університету імені Т.Шевченка: Економіка №85. – 2006, с.8-11.
7. Єріна А.М. Статистичне моделювання та прогнозування, Навчальний посібник. / А. М. Єріна – К.: КНЕУ, 2001. – 170 с.
8. Канторович Г.Г. Анализ временных рядов / Г.Г. Канторович // Экономический журнал ВШЭ. 2002. № 1. С. 85-16.
9. Канторович Г.Г. Анализ временных рядов / Г.Г. Канторович // Экономический

журнал ВШЭ. 2002. № 2. С. 251–273.

10. Литвинчук С.Ю. Информационные технологии в экономике. Анализ и прогнозирование временных рядов с помощью Excel: учебное пособие // С.Ю. Литвинчук – Нижегород. гос. архит.-строит. ун-т, Н.Новгород: ННГАСУ, 2010. – 78 с.

11. Лук'яненко І.Г., Жук В.М. Аналіз часових рядів. Частина перша: Побудова ARIMA, ARCH/GARCH моделей з використанням пакету E.Views 6.0. / Лук'яненко І.Г., Жук В.М. – К.: НАУКМА, 2013. – 188с.

12. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. / Магнус Я.Р., Катышев П.К., Пересецкий А.А. – М.: Дело, 2007. – 504 с.

13. Присенко Г.В., Равікович Є.І. Прогнозування економічних процесів: навч. посібник. / Присенко Г.В., Равікович Є.І. – Київ: КНЕУ, 2005. – 382 с.

14. Тихонов Э.Е. Методы прогнозирования в условиях рынка: учебное пособие. / Тихонов Э.Е. – Невинномысск, 2006. – 221 с.

15. Ширяев А.Н. Основы стохастической финансовой математики. Том 1. Факты. Модели. / А. Н. Ширяев – М., Фазис, 1998. – 512с.